

PRÉSENTATION DU NUMÉRO

À quoi sert la statistique publique ? Fondamentalement, elle sert à poser les bases d'une information de qualité à même d'éclairer de manière objective et rigoureuse le débat public.

Comment procéder pour y parvenir ?

Il faut déjà collecter les données élémentaires. À cette fin, le système statistique public réalise régulièrement des enquêtes sur le terrain. Certaines ont fait l'objet d'articles dans les précédents numéros du Courrier des statistiques, comme l'enquête Trajectoires et Origines [N8] ou l'enquête Emploi [N6]. De telles opérations sont nécessaires : elles permettent d'obtenir, parfois dans des conditions difficiles, des données qui ne sont présentes dans aucune base de données administrative, ni même privée.

Dans ce numéro, Pierre-Yves Cabannes et Thomas Lellouch nous présentent une enquête particulièrement emblématique, terriblement exigeante, qui fait l'honneur de la profession : l'enquête Sans domicile, menée au printemps 2025. Une telle opération, hors normes, ne peut en aucun cas reprendre une démarche standard et il a fallu concevoir pour elle une méthodologie sur mesure, tenant compte de la diversité des situations. L'originalité concerne de multiples aspects de l'enquête : le fait de passer par l'intermédiaire de services d'aide, la grande variabilité des configurations de terrain (hébergements, distributions de repas, accueils de jour et services itinérants), les « questionnaires qui parlent » en 21 langues, etc. La collecte a mobilisé près de 1 000 agents, pour un total de 17 000 entretiens. La complexité de sa mise en œuvre rend compte de la tension permanente entre exigence méthodologique et contraintes opérationnelles. Elle explique la rareté du dispositif, la dernière enquête équivalente datant de 2012. Dire que les résultats sont très attendus relève de l'euphémisme !

Pour nourrir le débat public, il faut par ailleurs s'appuyer sur une architecture solide permettant la comparabilité dans l'espace et le temps. Cela veut dire par exemple des concepts partagés, stables, éprouvés. Un autre pilier majeur d'un fonctionnement professionnel et maîtrisé, ce sont les nomenclatures. Discrètes mais puissantes, elles irriguent les statistiques publiques. On peut citer les nomenclatures socioprofessionnelles [N4], d'infractions [N7], mais aussi de catégories juridiques, de maladies...

Clotilde Masson nous parle ici d'une nomenclature jouant un rôle tout à fait majeur : la nomenclature d'activités. Sans elle, pas de secteurs ni de branches d'activité, qui sont centraux dans la grammaire de la statistique d'entreprise et de la comptabilité nationale. L'auteure décrit d'abord ce qu'est une nomenclature en général : c'est un objet riche, qui ne se limite pas à une liste de catégories, fût-elle structurée de façon arborescente. Celle des activités existe depuis fort longtemps, et elle a évolué au gré des transformations du monde qu'elle est censée représenter. Une nouvelle version de la nomenclature d'activités, c'est un événement très rare, qui a lieu tous les 10-15 ans, et qui s'appuie sur une préparation considérable. En particulier, depuis 1993, la nomenclature française est imbriquée dans celle de l'Europe, elle-même articulée avec celle des Nations Unies. Inévitablement, le résultat est fait de multiples compromis.

Si l'on veut que les statistiques publiques soient considérées comme une référence, il est nécessaire qu'elles inspirent confiance, qu'elles offrent des garanties de neutralité, de

transparence, de qualité. Cela requiert une capacité à se remettre en cause, et un cadre institutionnel solide : le Conseil national de l'information statistique (Cnis) [N6], le Comité du label [N5], l'Autorité de la statistique publique [N5], les revues par les pairs [N12].

Pierre Bischoff et Hervé Piffeteau montrent que cette confiance n'est pas un acquis immuable : elle se conquiert par le professionnalisme, mais un grain de sable peut la remettre en cause, et la rétablir est alors une conquête. Leur article est le deuxième volet d'une série chronologique de trois papiers sur le système statistique européen, dont le premier volet s'achevait en 2003. Notre grain de sable fait irruption en 2004 et c'est un gros grain : c'est la crise grecque, avec la falsification des comptes publics. Toute la statistique européenne s'en retrouve fragilisée. En raison de cette crise, mais pas seulement, l'Europe met en place une liste de principes (neutralité, transparence...) qui composent un « code de bonnes pratiques ». Celui-ci donne un cadre à l'ensemble des instituts nationaux de statistique (INS), qui demeure important aujourd'hui. Mais nous ne sommes pas au bout de nos peines, car un deuxième grain de sable, en l'occurrence une deuxième crise grecque, secoue l'édifice en 2009. Il conduit à consolider le code de bonnes pratiques, et notamment à s'assurer de sa bonne utilisation. Difficiles à vivre, ces crises n'en ont pas moins été fécondes pour consolider l'architecture institutionnelle des statistiques, et notamment sa législation par l'élaboration d'une nouvelle loi statistique européenne.

Produire des statistiques utiles au débat public, c'est aussi innover, en permanence, pour tenir compte des évolutions de l'environnement et du champ des possibles. L'innovation peut passer par de nouvelles sources de données [N13] : les données de caisse [N3], les comptes bancaires [N12], la téléphonie mobile [N13], les transactions bancaires [N13].

L'article de David Levy porte ainsi sur une source de données privées essentielle pour compléter les statistiques du tourisme : les données de plateformes (Airbnb, Booking, Expedia, Tripadvisor). Ici, l'innovation est aussi institutionnelle : après quelques expérimentations disparates, aux succès variables, c'est Eurostat qui a pris la main, jouant de fait un rôle de producteur de données utilisées par les INS, à front renversé. C'est d'autant plus légitime que ces plateformes de réservation ne se prêtent pas facilement à un découpage en pays : l'Europe semble la bonne maille. Envisager d'utiliser ces données nécessite toute une phase de préparation avec chacun des opérateurs, qui se matérialise in fine par des conventions. Les données de plateforme complètent très utilement le dispositif d'observation du tourisme sur un domaine difficile à appréhender : les locations de courte durée entre particuliers. Néanmoins, ce partenariat public-privé soulève des défis structurels : dépendance aux accords volontaires, opacité relative des algorithmes de gestion des plateformes et absence d'identifiant unique pour traiter les doublons de capacité d'hébergement. Pour l'avenir, un travail de consolidation s'impose.

L'innovation, nécessaire, ne se limite pas aux sources d'information. Elle renvoie aussi aux méthodes et environnements de travail des professionnels de la donnée. Dans le contexte de développement massif des traces numériques liée à l'activité, on assiste depuis quelques années à une évolution profonde, inspirée par les apports de la *data science* : nouvelles infrastructures comme le SSPCloud [N7], nouveaux formats [N9], processus plus standardisés d'intégration de données administratives [N9].

Lino Galiana et Olivier Lefebvre s'intéressent aux implications sur nos manières de travailler. Cette nouvelle discipline engendre progressivement une transformation en profondeur des métiers et l'émergence de nouveaux profils spécialisés : *data analysts*, *data scientists*, *data engineers*. La statistique publique ne peut ignorer ces transformations : elle s'attache à les mettre au service de ses missions, dans le respect de ses valeurs. La révolution de la *data science* favorise un rapprochement des compétences entre directions métiers et informatiques, au profit par exemple de la capacité à associer des données de formes et de natures diverses pour répondre à de nouveaux besoins, avec une exigence renforcée de transparence, un nouvel impératif de répliquabilité... Mais cela représente aussi un défi d'adaptation continue, dans un environnement où les besoins, les outils et les méthodes évoluent rapidement. L'Insee et les services statistiques ministériels se sont dotés d'infrastructures et d'organisations pour explorer, expérimenter, industrialiser et maintenir ces nouveaux processus.

Tout en adoptant ces nouvelles manières de travailler et en mobilisant ces nouvelles sources, on peut également innover institutionnellement, et aller au-delà du champ bien balisé de la statistique publique. Ainsi, depuis plusieurs années, des pôles de science des données, appelées aussi pôles *data*, se sont développés dans plusieurs inspections, par exemple à l'inspection générale des finances (IGF) [N13] et à l'inspection générale des affaires sociales (IGAS) [N13]. Leur fonction est différente, puisque leur rôle est avant tout d'aider à l'élaboration, au pilotage et à l'évaluation de politiques publiques sur des sujets souvent très ciblés.

Benjamin Belrhomari, Lauriane Martzel, Lisa Menez et Léo Zabrocki-Hallak nous parlent dans ce numéro de la *data science* à la Cour des comptes. Les juridictions financières – Cour des comptes et Chambres régionales et territoriales des comptes – ont depuis 2008, en plus de leurs missions historiques (contrôler, juger, certifier), une mission d'évaluation des politiques publiques. À cette fin, elles ont de plus en plus recours à l'analyse statistique comme élément de démonstration dans leurs rapports. Celle-ci leur permet d'aller plus loin dans leur appréciation du bon usage de l'argent public. Le département d'analyse et de science des données de la Cour des comptes, de par le positionnement de celle-ci, possède la particularité de pouvoir accéder à de très nombreuses sources administratives. C'est une équipe conséquente de *data scientists* et *data analysts*, ayant déjà plusieurs années d'expérience. Elle emprunte son cadre conceptuel et ses méthodes à la recherche tout en conservant l'essentiel de ses propres normes et procédures. Comme c'est le cas pour les pôles *data* des inspections, elle est soumise à une exigence de rapidité, qui modifie la manière d'appréhender les sujets par rapport aux démarches de la statistique publique.

Enfin, nourrir le débat public, c'est aussi s'inscrire dans le temps long, comme on a pu le voir avec l'histoire du lien entre statistique publique et débat démocratique depuis la naissance de l'Insee [N9 et N11], celle de la statistique européenne depuis l'origine [N14], ou celle des échanges de données et publications statistiques depuis le XIX^e siècle [N14].

Béatrice Touchelay remonte ici au XIX^e siècle pour s'intéresser à un sujet parfois oublié, celui de la statistique dite « coloniale ». Les archives, et en particulier certaines déposées très récemment, apportent un éclairage original sur la manière de réaliser des statistiques au temps de l'empire colonial : on y trouve des lettres de responsables mais aussi

d'enquêteurs, des tableaux de suivi, des parcours de collecte... On constate, avant la Seconde Guerre mondiale, la difficulté à mettre en place une organisation adéquate dans les territoires pour produire des statistiques harmonisées, faute de moyens et de compétences, voire de sensibilité aux enjeux. On voit aussi à quel point les concepts, les nomenclatures, se fondent implicitement sur une vision du pays colonisateur pouvant se révéler très décalée par rapport à la réalité du terrain.

Ce numéro N15 met ainsi en lumière de multiples facettes d'une profession qui a vocation à rendre compte de façon objective et aussi rigoureuse que possible du monde qui l'entoure : grandes enquêtes, grandes nomenclatures, contexte institutionnel européen, nouvelles sources de données, nouvelles démarches et environnements de travail, entités voisines de la statistique publique, et pour finir une description historique nous apportant en creux un recul utile sur les exigences et limites du métier.

Bonne lecture !

Pascal Rivière
Directeur de la collection, Insee
